DOCUMENT RESUME

ED 121 860 TM 005 297

AUTHOR	Mushkin, Selma J.; Billings, Bradley B.
TITLE	Generating Outcome Measurements: Achievement and Attitudes. A Guide to Educational Outcome Measurements and Their Uses. Seminar No. 3.
INSTITUTION	Centro Multinacional de Investigacion Educativa, San Jose (Costa Rica).; Georgetown Univ., Washington, D.C. Public Services Lab.
SPONS AGENCY	Agency for International Development (Dept. of State), Washington, D.C.
PUB DATE	Nov 75
NOTE	35p.; For related documents, see TM 005 295, 296 and 298-300
EDRS PRICE	MF-$0.83 HC-$2.06 Plus Postage
DESCRIPTORS	Academic Achievement; Affective Tests; Comparative Education; *Educational Assessment; Educational Planning; *Guides; *Measurement; *Measurement Techniques; National Surveys; Program Development; Seminars; Standardized Tests; Student Attitudes; Student Evaluation; Student Testing; Test Construction; Testing Problems; *Testing Programs
IDENTIFIERS	*Educational Outcomes

ABSTRACT

This guide is essentially designed as a teaching aid for those who would inform planners, officials of educational ministries, school administrators, principals, and teachers about educational outcome measurements. In outline and graphic form, the guide presents topics for discussion in a seminar dealing with how to obtain information on educational achievements and attitudes as outcomes. An outline is presented and is intended as a preliminary "lesson plan." A discussion leader, selected by the group, will select materials appropriate for his or her group, adding and illustrating materials that are of particular interest to his or her country's (or region's) educational system. Steps required to use the seminar format are listed. At the end of the seminar those participating should: (1) Understand that specific choices have to be made about selecting proxies or indicators of several types of outcome measures that come as close as is feasible to capturing the defined purposes and providing a reasonably accurate and valid count; (2) Understand what some of the major classes of choices about outcome measurement are in terms of counting knowledge, skill, and attitudes; (3) Be familiar with general approaches to obtaining the information required; and (4) Be able to enumerate the major arguments in favor of or opposed to major options in obtaining the information required. (RC)

Documents acquired by ERIC include many informal unpublished materials not available from other sources. ERIC makes every effort to obtain the best copy available. Nevertheless, items of marginal reproducibility are often encountered and this affects the quality of the microfiche and hardcopy reproductions ERIC makes available via the ERIC Document Reproduction Service (EDRS). EDRS is not responsible for the quality of the original document. Reproductions supplied by EDRS are the best that can be made from the original.

SEMINAR NO. 3

GENERATING OUTCOMES FOR ACHIEVEMENT AND ACTIVITIES

A

U.S. DEPARTMENT OF HEALTH
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY.

Educational Outcome Measurements and Their Uses

TM005 997

A GUIDE TO

EDUCATIONAL OUTCOME MEASUREMENT AND THEIR USES

SEMINAR NO. I
Types of Outcome Measurements

SEMINAR NO. II
Uses of Outcome Measurement

SEMINAR NO. III
Generating Outcome Measurements: Achievement and Attitudes

SEMINAR NO. IV
Generating Outcome Measurements: Economic and Societal

SEMINAR NO. V
Applying Outcome Measurements

SEMINAR NO. VI
Feedback Consequences and Steps toward Implementation

EMINAR NO. III

GENERATING OUTCOME MEASUREMENTS: ACHIEVEMENT AND ATTITUDES

Selma J. Mushkin, *Director*
Bradley B. Billings, *Research Associate*

PUBLIC SERVICES LABORATORY
Georgetown University
CEMIE (Centro Multinacional de Investigación Educativa)
Costa Rica
November, 1975

This series was prepared as part of a project sponsored and funded under a contract with the Agency for International Development. The views expressed are those of the authors and not necessarily those of the Agency.

PREFACE

This guide essentially is designed as a teaching aid for those who would inform planners, officials of educational ministries, school administrators, principals and teachers about educational outcome measurements.

In recent years, educational services and facilities have made increasing claims on national economic systems. Since the needs of all public services, including education, are pressing and sum to totals far in excess of resources available, it has become urgent that governments initiate processes of questioning both existing resource uses and those proposed, asking: Can we serve the public better and more cheaply? What can we do that would more surely achieve the results we seek?

Responsible education, planning, and finance officials have indicated great interest in measurement of educational results as a beginning step to finding answers to these questions.

In outline and graphic form, this guide presents topics for discussion in seminars dealing with the following subjects:

I. Types of outcome measurements
II. Uses of outcome measurement
III. Generating outcome measurements (achievements and attitudes)
IV. Generating outcome measurements (economic and societal)
V. Applying outcome measurements
VI. Feedback consequences and steps toward implementation.

The outlines for six seminars which follow are intended as guides or preliminary "lesson plans." The discussion leader will select material appropriate for his group. His additions of information and illustra-

tive materials that are of particular importance to his country's (or region's) educational system would greatly enhance the discussion.

"Figures" are presented before each discussion question and are tended to serve as visual aids. The discussion leader may distribute copies to group members or he may find it helpful to enlarge the figures into charts to use for the discussion, or to have view graphs made for this purpose. To facilitate such use all figures are reproduced in appendix B.

The topics covered can be so selected that a general orientation about educational outcomes may be presented in summary form in a single seminar, or more in depth discussion would require at least 6 seminars. Extensive training of one year or more is required to develop skills needed to properly design measurement instruments or the analytical capacity to apply them. However, for those who have the required analytical and quantitative skills already, participation in the seminars outlined would help provide an understanding of the policy implications of the emphasis on outcomes. The seminars are not designed as technical training, rather they are intended to:

- Provide an understanding of the different types of educational outcomes.
- Identify the variety of uses of outcome measurements.
- Help convey the choices on outcome measurement instruments and yardsticks for policy officials, school administrators and teachers.
- Provide some rudimentary understanding of the tools of analyzing the correlates of education outcomes.
- Make plain the possible impact of emphasis on educational outcomes for educational planning and school finance, as well as work in the classroom.

To use the seminar format the following steps are required:

STEP 1
A seminar discussion leader must be selected.

STEP 2
The discussion leader, together with the educational officials sponsoring these sessions should identify the particular purposes of the seminar, and should answer such questions as: For whom are the seminars intended? What is the purpose of carrying out the seminar?

STEP 3
Seminars should be announced as an offering with (a) the clear indication of top level support, and (b) a practical use of the learning

achieved in the continued work of the official, administrator, or teacher.

STEP 4

The discussion leader might select from the materials and "figures" presented in this guide those that appear particularly useful for the specific groups with whom the discussions are scheduled.

As an aid in this selection process, Appendix A suggests some portions of the text that appear most important for the following groups:

- top officials concerned with overall policy planning
- middle management in Ministries of Education and Finance
- school administrators
- school principals
- teachers
- instructors in teacher training institutions.

STEP 5

Supplementary reading materials should be selected. Supplementary readings are available from the Public Services Laboratory of Georgetown University: these include **Educational Outcome Measurement in Developing Countries: An Annotated Bibliography;** and **Educational Outcome Measurement in Developing Countries.**

STEP 6

Some preliminary evaluation should be introduced at the close of the seminars to determine whether the objectives of each seminar have been fulfilled. This can be done by testing instruments, by questionnaire, or by final closing comments of participants.

Such evaluations tell us little about whether the seminars made a difference. Only practical application indicate whether the seminars were successful in changing traditional practices.

TOPIC TO BE COVERED:
HOW TO OBTAIN INFORMATION ON EDUCATIONAL ACHIEVEMENTS AND ATTITUDES AS OUTCOMES

Objectives of Seminar No. III

At the end of the seminar, those participating should:

(1) Understand that specific choices have to be made about selecting proxies or indicators of several types of educational outcome measures that come as close as is feasible to capturing the defined purposes and providing a reasonably accurate and valid "count."
(2) Understand what some of the major classes of choices about educational outcome measurement are in terms of counting "knowledge," "skill," and "attitudes."
(3) Be familiar with general approaches to obtaining the information required.
(4) Be able to enumerate the major arguments in favor of or opposed to major options in obtaining the information required.

Introductory Question for Discussion

What data are now available that could be used in measuring educational achievements and attitudes? What new data have to be collected?

NOTES FOR DISCUSSION LEADER

The discussion leader could open the discussion by showing that the measurement of educational outcomes in developmental terms requires assessment of the contribution of education to development-economic, social and political. This contribution describes the functioning of the adult member of society. It could also be stressed that there are other measures of outcome that help define the product of the educational establishment starting with measurements useful for judging the knowledge, skill, and attitudes of young children and school leavers. It is important that this period not be neglected; a good proportion of a person's lifetime is the growing years. And where years of compulsory schooling are long, a large share of a lifetime is represented by the years while at school.

Education has the purpose of transmitting knowledge but it also has a broader and deeper purpose. It is the whole of the person that is the target of education—his knowledge, skill, attitudes, creativity, and other attributes. The variety of aspects need to be measured, lest education narrowly viewed breeds men with narrow vision.

Participants could be asked to name some existing sources for data that could be used in measuring educational outcomes. Examples might be standardized or teacher tests, information about school leavers (years of schooling, their employment) and teachers' descriptive records of student achievement, skills and attitudes.

Ideas on Presentation

For young children, such outcomes as employment and earnings or political participation are activities of the future. Their immediate learning and competence can be recorded, however, by measurement of knowledge and skill achievements, and by measurements of attitudes such as those toward society by, for example, scales on internal and external control, and by measurement of attributes such as self-esteem and creativity. Similar measures of knowledge, skills, and attitudes of adults also are important to give educational planners the information they require to assess the need for subsequent action. To judge the outcomes of the schools, it is possible to assess the competence of those who have left school, but this alone tells us little about the results of the processes of education—for example, about the impact of textbooks on the knowledge acquired.

Would it make a difference if the children in the first and second primary grades had books to use? Would it make a difference if the teacher had a secondary education? Would it make a difference if TV were used to inform and build new competencies in adults in the community? Such questions of outcome of education can only be answered in terms of the learning achieved. And for these purposes, we need measures of knowledge, skills, attitudes, and attributes.

The discussion leader can conclude by emphasizing that outcome measurement is essentially a matter of measurement of what individuals know or what they have learned, that is, of testing the knowledge, skills, attitudes, and attributes acquired. Such testing is especially important for children because substitute measures are harder to develop. The leader can then set the seminars in perspective by showing that in this seminar we look at the various measurements of outcomes. We start by exploring options of how to implement achievement and attitude testing. In a following seminar, we examine economic, social, and cultural outcomes linked to current economic development.

Questions for Discussion

Are uniform tests useful for educational planning now being given to those in school or school completers? Are new instruments and methods needed to judge knowledge, skills, and attitudes of the population, and those in school?

NOTES FOR DISCUSSION LEADER

The formal school system uses many tests; mostly these are tests developed by the teachers for their own use. However, school systems in developing countries often also have uniform examinations at terminal points in the educational system, for example, at Grade Six, or later, as evidence of completion of the course of high school study and eligibility for the university.

Mostly these examinations in the past have been patterned on the English or French educational systems, making the diploma holder eligible for entry into a British university or one in France. Moves are underway in many developing countries to break away from these traditional links and to formulate different examinations more suitable to the purposes of the nation. Arguments are being advanced on both sides of the issue.

Throughout this discussion it would be useful to remind the participants that the decision on selection of what tests to use requires a careful assessment of the immediate policy applications, the adaptations required to match testing program with learning objectives of the schools (or educational ministry) and the relative costs of such adaptations versus development of new testing instruments. Questions of functionality patently are important because the knowledge, skills, and attitudes sought through education must be functional to the social, economic, political, and cultural characteristics of the nation.

An understanding of the skills and knowledge of the population—those in school and those not in school—is fundamental information for educational planning and decision about use of resources for education. For those in school, there is easy access to the student and perhaps to a testing program that, if carefully developed, can quantify the knowledge and skills learned.

For those out of school, understanding about the levels of competency is harder. But the assumption may be made that by observation in rural communities, the competencies in the population, even by some approximate age grouping, can be achieved.

In carrying out the discussion on options for testing educational achievement, it might be useful to just ask the group to identify the uniform tests now being given. Discussion of current testing might be broadened to include an exchange among participants on the purposes of such testing programs. In concluding the discussion, the conversation might be directed to the application of such testing programs in rural areas and of possible substitutes for "pad and pencil" testing.

Ideas on Presentation

Nations that seek to measure cognitive achievement of children and adults or to measure certain other aspects of competence such as attitudes or attributes face a number of options. Figure III-1 lists some of the possible approaches to assessment.

• **Constructing new national tests:** An important option in judging knowledge and skills is to develop national tests, as San Salvador, Thailand, and Indonesia have done. This method of providing testing instruments looks to a committee or agency designated by the chief education officer as responsible for test development. In the Philippines, for example, a university grant committee was established to develop tests for university admission. A national testing instrument may require a special administrative unit to act as the test developer and administrator.

It is possible to permit different school systems and schools to develop their own tests designed to match the specifications of local school officials or a principal of a particular school. However, unless there is superimposed an anchor test on these divergent testing instruments, the diversity would continue to bar comparative analysis of educational outcome in relation to resources used. (The anchor test would be designed essentially to provide a method of converting results of one test to an "anchor" or standard finding.)

Among the advantages of developing new national tests are that they: (1) can offset effects of cultural influences on test performance; (2) can meet specific curriculum and educational objectives of school system; (3) can supply particular outcome data required by planners; and (4) provide opportunity to increase awareness and expertise with educational measurement among teachers, administrators, and educational researchers.

The disadvantages include: (1) the process is time-consuming; (2) developing original testing instruments is far more costly than adopting existing instruments; (3) skilled manpower for the process may be in short supply and more critically needed in other education areas. These can be partially overcome through the development of regional testing centers serving several countries with similar educational problems and goals.

• **Tests developed in other countries:** Another option is to use the testing instruments developed essentially in other countries. A number of countries in the Caribbean, for example, administered

FIGURE III-1

OPTIONS ON TESTING EDUCATIONAL ACHIEVEMENTS

I. Standardized Tests

- Own national test development
- Tests developed by other nations with same language
  —with same cultural patterns
  —with different cultural patterns
- Tests developed in other nations with different language
- Tests developed for international comparisons
- Regional tests developed by OAS or other international agencies
- Local school system tests with an anchor test for comparative outcome analysis

II. Teacher-made Tests

- Use of locally designed teacher-made tests

III. Surveys and Observational Reporting

- Surveys of those participating in program
- Recorded observations of community behavior patterns

the leaving examinations of the British system, although steps are being taken to develop an Island-wide testing program. Some of the French speaking countries use the Baccalaureat and sixth grade entrance examination of the French system.

These examinations have the disadvantage of being adapted to the objectives of cultures that make the tests; they may be neither

relevant nor reflective of objectives in countries with different cultures. Moreover, such examinations create barriers to educational opportunities. For those nations whose university education program is tied to the United Kingdom or France, the use of examinations of those countries serve, however, to facilitate entrance to the foreign universities.

The examinations of these kinds are essentially intended to "screen out" those judged not to be qualified for higher level education. But testing for purposes of understanding about existing levels of competence in a nation or planning current school services serves very different purposes and requires different instruments and procedures.

• **Tests developed for international comparison and regional tests:** Another choice available to nations seeking standardized testing instruments for outcome assessment is the use of tests developed for international comparisons. Most prominent among these are those developed by the International Association for the Evaluation of Educational Achievement (IEA), which began a cross national survey on six subject areas in 1966. A major difficulty with this choice for a developing nation is that the primary objective of the IEA survey was rather different from that of a nation conducting outcome assessment. And many of the competencies tested may not be functional for a particular nation.

It has some advantages, however, among these is the availability of the subject matter tests in Spanish for inter-country comparisons.

OAS has for some years been developing materials related to educational outcomes and their measurement. In 1972, the first Inter-American Seminar on Educational Research was held in Costa Rica. That Seminar considered the relevance of multi-dimensional outcome measurement and application in evaluation, including measurements of interrelation of measures of intelligence and attitudes, and particularly of attitudes toward special target groups. Factor analysis of background variables was presented as methodology for assessing determinants of learning. In 1974, OAS developed instrumentation for use in evaluation, particularly of the OAS financed Educational Television projects.

• **Teacher-made tests and classroom evaluation:** Another source of educational outcome data is the local school classroom.

Standardized testing instruments are not always available and are not always appropriate for all purposes. At the classroom level,

clearly evaluation is needed throughout the period of instruction, not merely at the end. Tests written by teachers are well suited for this periodic measurement of outcomes. They serve a number of functions:

(1) Assist the teacher in evaluating the adequacy of instructional techniques.

(2) Assess student progress.

(3) Motivate and focus learning.

(4) Provide a basis for a course grade.

The course grade given by teachers is generally influenced by other factors than test scores. Teachers must rely on observation and recordkeeping to evaluate important student behavior such as initiative, self-direction, curiosity, creativity, leadership skills, organization of materials, use of the library.

• **Observation and surveys of functional competence:** By survey instruments and observation, the competency of those in the community can be recorded. For example, if an instructional program is introduced to provide information about farming and personal hygiene, the knowledge and skills required can be measured by

- testing the participants to determine the knowledge that they have acquired
- querying the participants about their views on what they studied
- observing the participants to determine the practices they follow.

Differences can be assessed between what was known and practiced in the community before the program and after the program. More particularly, differences between communities selected on a random basis to provide experimental controls as well as experience with instruction could provide information about the consequences with respect to outcomes but the outcome measurement itself still requires analysis and selection.

Surveys may yield required information in some instances; in others, it might be preferable to station temporary (one week, for example) observers to record based on their observations such things as methods of farming—use of fertilizer, tools, pesticides or methods of personal hygiene—hand washing, care of infants, food preparation, and so forth.

TEST AND SURVEY FORMULATION

Question for Discussion

What has been the experience with test or survey formulation?

NOTES FOR DISCUSSION LEADER

Development of testing or survey instruments is expensive and time consuming. Before a testing or survey instrument is to be used, a substantial period should be allowed for development and initial trials of the instrument. Pretesting of survey or test instruments may suggest inadequacies and the instrument may have to be revised in the light of the initial trial.

In general, at least four to six months are required to formulate a schedule questionnaire or simple test and to pretest, validate, finalize, and duplicate it. Any timetable that assumes staff, except a very sizeable one, to construct a new text and validate it in a few weeks would create difficulties.

Figure III-2 shows the steps in construction of a uniform testing instrument. Survey instruments also require as elaborate a set of procedures.

It is important that the costs of developing new tests or new survey instruments be fully accounted for so that the decision on adaptation of existing instruments or development of new ones be better informed. Steps involved in adaptation of existing instruments are enumerated in a subsequent presentation of Figure III-3.

It is important once a decision to develop a new test has been made to consider a number of additional questions and especially the questions: What is the testing instrument (or survey instrument) intended to do? What is the test measuring?

Ideas on Presentation

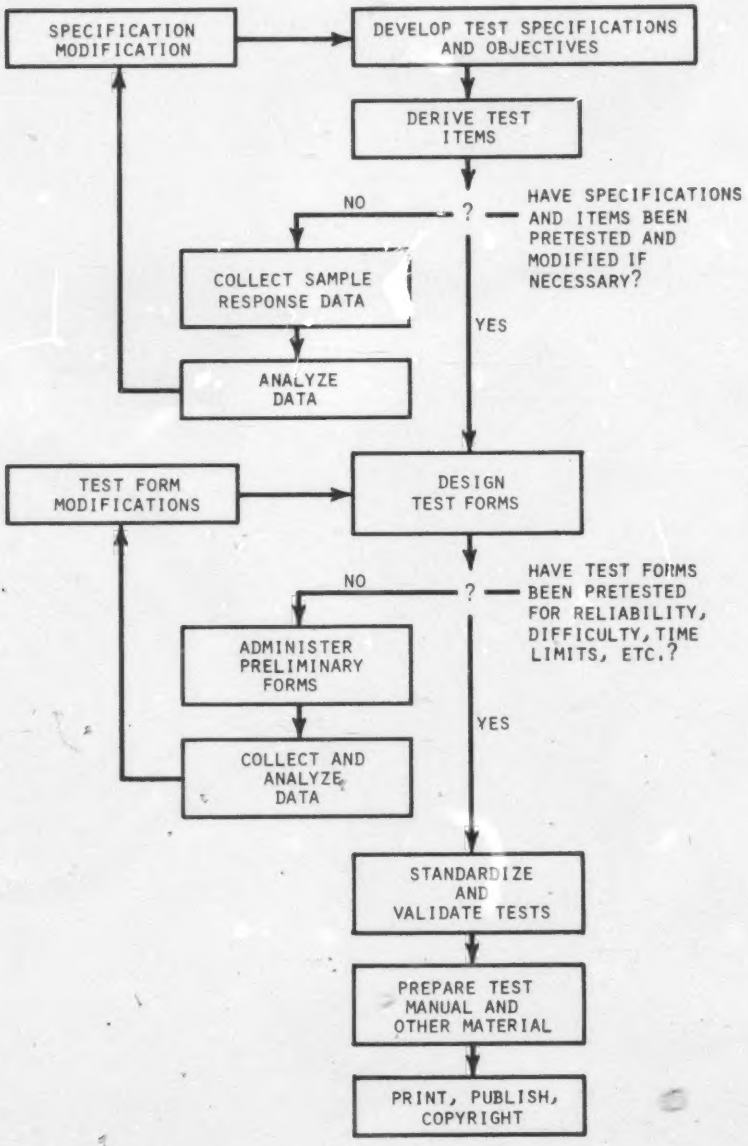
Figure III-2 shows the steps involved in construction of a test. Having decided on a measurement program for policy analysis evaluation, a major step is identification of the objectives of the learning being tested. What is it that the test is testing? What information, knowledge, skill is essential to an understanding and application of the "bit" of knowledge or skill identified?

Often teachers, teacher training institutions, school administrators, as well as educational planners, are asked to participate in the compiling of a set of questions that would capture at least on a proxy basis the range of knowledge that it is intended that those completing some designated grade level know or understand.

The derivation of the testing items requires initial collection of sample response data for analysis. Following this analysis and any

FIGURE III-2

STEPS IN CONSTRUCTION OF TESTS



revisions suggested, a formal pretrial might well be initiated to help assure that the tests are validated. The validated tests would then be prepared for distribution.

Among the questions that have to be answered about the testing instruments are these:

• What is the primary purpose of the testing program?

Whether it is national assessment, individual assessment for selection purposes, guidance for students and teachers, or some combination of these, the purpose of the testing program affects many of the other problems listed below.

• What goals and objectives of the educational system are to be assessed in the testing program?

The educational objectives and the weight attached to each determine the content of assessment instruments.

• Which grades or levels will be tested?

To assess educational outcomes, it is not necessary to test all grades. But if the assessment is to have useful feedback effects for students and teachers, assessment should be done periodically as a person progresses through school, not merely at its conclusion.

• Will all students be tested or merely a sample?

A sample is sufficient for the purpose of measuring national accomplishment of educational goals. But for direct benefits for individual students and teachers, all students should be tested.

• Will norm-referenced or criterion-referenced instruments be used?

Norm-referenced tests show how students rank in relation to each other. Criterion reference tests are designed to show specifically what a child or group of children knows and does not know. For example, a criterion referenced mathematics test may reveal that the student has a grasp of fractions but cannot perform long division. The norm referenced test might merely show that he performs better than 60 percent of his classmates.

ADAPTING EXISTING TESTS AND SURVEY INSTRUMENTS

Question for Discussion

What has been the experience with adapting existing testing instruments?

NOTES FOR DISCUSSION LEADER

Adapting existing instruments may be far less costly than an entire new instrument development. But the group would want to consider the steps involved in adjustment of existing instruments and the costs that would be incurred in carrying out the work involved.

Experience of countries that have adapted instruments is extremely helpful. But this experience would have to be probed to determine, among other things, how the adapted instrument was applied. Also to be determined are such issues as:

- What the full costs were of the adaptations?
- What weaknesses were found in the course of application?
- What additional steps are needed to usefully adapt an existing instrument?
- What steps were taken to validate the adapted instrument?

Ideas on Presentation

An important option to be considered is to determine whether existing test instruments designed in developed countries should be adapted and if so, which ones? This option has the advantage of the relative speed and economy of adaptation compared with new test construction and validation.

The procedure of adapting a test developed in and for a different culture includes many necessary steps:

—**Translation** is needed. Vernacular phrases or proverbs must be rendered understandable in the new language.

—**Content adjustment** often becomes necessary. On a superficial level, content changes may involve changing monetary or measurement systems; more subtle changes may need to be made, reflecting differences in cultural traits or work habits.

—**Testing mechanics** often pose problems. Successful use of existing testing instruments implies conditioning the test takers to standardized testing and procedures. For example, many practitioners in this area have found that separate answer sheets cause confusion and hence waste time—despite the inconvenience, answers within the

FIGURE III-3

LEVELS OF ADJUSTMENT FOR ADAPTATION OF EXAMINATIONS AND SURVEY INSTRUMENTS

(1) Translation of language.

(2) Changing the vocabulary for the proper reading level.

(3) Changing illustrations and photographs to reflect local conditions

(4) Adjusting procedures implied or specified to match the expectations and experience of the learners.

(5) Adjusting the content to reflect the local culture and life style.

(6) Accommodating the learning styles of the students.

test booklet itself are preferable. Another example: cross-cultural research shows that people have very different senses of time—and time limitations may have to be adjusted.

—**The test rationale** may not be suitable. Every test is designed within a particular cultural context, and is usually validated against practical criteria within that particular culture. Poor facility in language, slow work habits, or lack of abstract thinking will probably affect test scores. The purpose of the test will dictate whether these characteristics should be captured by the test.

Figure III-3 shows six levels of adjustment necessary for adaptation of instructional materials and examinations. Experience by Haile Sellassie I University in adapting testing instruments shows that the last three steps are complex and require expert staff with intimate knowledge of the culture. Finally, the process of validation (insuring accuracy) must be carried out in the adapting country—validation in the country from which the test was taken does not insure accuracy.

TEACHER TESTS AND EVALUATIONS

Question for Discussion

In what ways does it seem feasible and equitable to use teachers' evaluations to assess educational outcomes?

NOTES FOR DISCUSSION LEADER

Most discussants would argue that the variation in tests from one teacher to another provides insufficient comparability and, accordingly, they are not very useful for overall planning purposes.

While there are limitations in the use of teachers' tests, they serve important instructional purposes permitting periodic assessment of what the students have learned of a given curriculum and what remains to be learned. If accompanied by records of behavior that are for reporting purposes and not for grading or judging, it may be possible to overcome some of the subjectivity in evaluations not tied to objective test questions.

In some instances, it may be feasible starting with a teacher's test to gain collaboration among teachers in a given grade and subject matter in developing a uniform test for that grade and subject. Some discussion of such an effort would give greater specificity to problems of developing and applying uniform tests through a nation. Also, after analyses that provide teacher test adjustment factors, it may be feasible to use such grades for planning purposes.

Figure III-4 compares characteristics of standard tests and teacher evaluation according to three criteria, namely, validity, content appropriateness and uses. In addition, other advantages and disadvantages of each form of test are outlined.

Participants might be asked to review the advantages and disadvantages enumerated and to give their views on these.

Ideas on Presentation

If standardized examinations are not available, teacher grades can provide some information toward a widescale outcome assessment. However, adjustments must be made to compensate for the subjectivity of the method of the widely varying situations in schools. To develop such an adjustment scheme, a set of indicators to measure school differences is needed. Some possible indicators are:

(1) differing school policies on grading; and

(2) rating schools on

(a) number of students going into higher educational institutions;

(b) number of former students employed in various job categories (or not employed); and

(c) general behavior of school leavers—reading of newspapers, or use of other communication methods such as radio and attendance at plays, concerts, etc., participation in professional or other organizations, charitable pursuits, and church activities.

FIGURE III-4

RELATIVE CHARACTERISTICS OF STANDARDIZED TEST

	Validity	*Content Appropriateness*
Standardized Tests	Statistical and empirical evidence of validity	keyed to national or regional goals and priorities
Teacher Evaluation	no direct evidence of validity	keyed to instructional unit in the classroom

ND TEACHER EVALUATION

Uses	*Advantages*	*Disadvantages*
• planning • comparative outcome evaluation of regions and schools • prediction of later school success • feedback to teachers on knowledge of students • feedback of nonformal education to formal for placement, for credentials, etc.	• evaluate all participants on the same set of criteria • standardized tests used at intervals through school career provide longitudinal information • evaluate knowledge, skills where there are no teachers (nonformal education; informal education)	• provide limited information about a student's skills and knowledge • may control instruction and course content • may not measure knowledge covered in local schools
• comparative outcome evaluation of individual students • feedback to teachers on success of instructional methods	• grades are based on cumulative record of performance • show typical behavior and performance in daily classroom situation • provide teacher and student with periodic assessment of achievement • may include assessment of behavior that cannot be measured by tests of skill knowledge	• short-range cognitive and affective changes are not recorded over the school career; cumulative records may not substitute for longitudinal records built up by standardized tests • often compare students within a class instead of providing data on how much has been learned • may be more influenced by good conduct, docility, neatness than on how much a student knows • pressure from parent or community may influence grades

Questions for Discussion

Is there agreement with a measurement specialist who urged that "self-esteem" and "external-internal controls" were especially important noncognitive capacities to be developed through education? Or are attitudes about work—especially manual labor—a more significant outcome for developing countries?

NOTES FOR DISCUSSION LEADER

Assessment of attitudes and attributes of individuals poses special problems for the educator and the planner. For one thing, attitudes and attributes clearly reflect the multitude of experience of the individual—in the home, in the work place, in the community, in the church. They also are not uniformly defined. There is far less agreement on the purposes of the educational policies regarding attitudes and attributes than on cognitive skills. When schools do make a difference, the attitudinal impact sometimes works as a change agent, as has been emphasized in the UNESCO publication, *Learning to Be*.* Schools, noted the author, Faure, sometimes inculcate values into school children that estrange them from their surroundings and push them out of rural areas into urban settings. And in some countries such as Tanzania, deliberate efforts are being made to overcome this attitude.

To illustrate the range of measurements, a list developed by an Arab educator includes the following values expected from education.

resourcefulness	*adaptability*	*responsibility*	
reality	*efficiency*	*enthusiasm*	*sensitivity*

If attitudes and attributes are measured, decisions become necessary on four issues:

(a) How are the most significant variables to be selected among the list of variables?
(b) What instruments are to be used to do the measurement?
(c) What items and how many are to be selected from these instruments?
(d) Are new instruments to be developed or existing instruments to be adapted for use?

Several guides have been developed of measurements of attitudes and attributes. It may be useful to use these guides when a decision has been taken to measure noncognitive qualities. Two constructs in particular—self-concept and locus of control—are seen as important to understanding

* Faure, Edgar, *et al. Learning to Be. The World of Education Today and Tomorrow*. Paris, France: Harrap, UNESCO, 1972.

competence. There are many testing instruments for each construct but especially numerous are the instruments measuring self-concept and the related construct "self-esteem." It becomes essential to make a choice among instruments and also for any particular instrument to selectively choose those items that are appropriate to the country or region in which the testing is being done.

The types of questions raised earlier about adaptation in connection with achievement testing apply here as well except attitudes are in many ways less objective, more personal. Certainly, cultural patterns, sensitivity to question, likely concerns of respondent about use of the information, politically and economically, all have to be considered. The way in which the question is asked will be important as well as the person who asks it, especially in places where illiteracy makes "pad and pencil" testing impractical.

Ideas on Presentation

In addition to academic skills and knowledge, a testing program should include measurement of those social and motivational skills considered important for national development. For example, two important variables for which instruments are readily available are self-concept and locus of control. These are being experimentally applied in both Tunisia and Ethiopia.

The 1966 Report on Equality of Educational Opportunity* found, on the basis of six self-concept and locus of control questions (see Figure III-5A) included in its survey of almost half a million school children, that the attitudinal variables accounted for more of the variation in achievement than any other set of variables.

Self-concept, or one's view of oneself, is a critical personality variable guiding much of the individual's future development and accomplishment. **Locus of control** (sometimes called "internal-external" control) refers to a person's view of events affecting his or her life. An "internal" person feels in control of these events; an "external" person feels they are controlled by fate or chance. It has been found that the "internal" attitude is more likely to contribute to modernization and economic development.

If these and similar psychological constructs are important to learning, a host of questions arise about critical timeframes in which attitudes can be changed.

For cognitive skills, it is generally assumed that almost all persons

* Coleman, James S. *et al. Equality of Educational Opportunity*. Washington, D.C.: U.S. Department of Health, Education, and Welfare, U.S. Government Printing Office, 1966.

can acquire the "knowledge." For some, the process of learning may be longer than for others but the learning will take place. If, however, it is not assumed that cognitive learning is essentially dependent up learning to honor oneself or to have self-esteem, then the factors determining learning are broader and the achievement of learning less certain.

Items on attitude toward work are included in Figure III-5B. The items for preference scales were developed in Tanzania for purposes of evaluating their economic projects in schools. The validity responses are heavily dependent upon the way in which the Tanzanians view the inquiry and inquiries. It is easy to imagine that in some places the answers to the questions would be intentionally constructed to please the questioner.

FIGURE III-5A

SELF-CONCEPT AND LOCUS OF CONTROL QUESTIONS*

(1) **How bright do you think you are in comparison with the other students in your grade?**

Agree or disagree:

(2) *"I sometimes feel that I just can't learn."*

(3) *"I would do better in school work if teachers didn't go so fast."*

(4) *"Good luck is more important than hard work for success."*

(5) *"Every time I try to get ahead, something or somebody stops me."*

(6) *"People like me don't have much of a chance to be successful in life."*

Answered either like me or unlike me:

(7) *"I often wish I were someone else."*

(8) *"If I have something to say, I usually say it."*

(9) *"I often get discouraged at what I am doing."*

*The first six questions are from Coleman Report (see Seminar III) and the remaining three are additional sample test items on self-esteem.

FIGURE III-5B

ITEMS ON STUDENT ATTITUDES TOWARD FARM WORK

- *"I like working on the farm because I enjoy it."*
- *"I do not mind doing gardening if its earns me money."*
- *"I like gardening as a hobby."*
- *"I like working on the farm because I have company there."*
- *"If I could choose I would do more interesting things than farming."*

SOURCE: *African Regional Seminar on Education Evaluation.* International Institute for Educational Planning and International Bank for Reconstruction and Development. Dar es Salaam, Tanzania: April/May 1975.

Selected Uses

As indicated in the Preface, some of the materials and "figures" in the six seminars may be particularly useful for the specific groups with whom the discussion are scheduled.

To some extent, selections will be made at the discretion of the discussion leader and may depend in part on the length of time available and on the amount of material included in the seminar.

Brief Comment on Selections from Seminars for Different Groups

Generally, Seminars III and VI, along with introductory material from Seminar I, are the most important for teachers, officials directly concerned with school administration, and teacher training institute instructors.

Seminars II, IV, and VI are most important in providing basic materials for top officials concerned with overall policy planning, middle management officials in Ministries of Education and Finance, and training institute instructors who teach planners and school administrators. Seminar I provides a backdrop discussion of concepts and of the multiplicity of outcome measurements. By and large, Seminar V will help inform top officials in planning and in Ministries of Education and Finance of what they can expect from the analyses.

Selected Uses for Seminar III

All materials in Seminar III should be presented to school administrators, principals and teachers in both schools and teacher training institutions. If the seminar is given for teachers only, stronger emphasis could be given to the last two discussion topics:

pp. 19-21 Teacher Tests and Evaluations
pp. 22-25 Attitudes and Attributes

For administrators and principals, emphasis could be given to:

pp. 9-13 Options on Testing for Educational Achievements
pp. 14-16 Test and Survey Formulation
pp. 17-18 Adapting Existing Tests and Survey Instruments.

Top planning officials and middle management officials in the Ministries of Education and Finance would benefit from a very general overview of the seminar and presentation of the first two topics:

pp. 7-8 Introductory Question
pp. 9-13 Options on Testing for Educational Achievements.

FIGURE III-1

OPTIONS ON TESTING EDUCATIONAL ACHIEVEMENTS

I. Standardized Tests

- Own national test development
- Tests developed by other nations with same language
  - —with same cultural patterns
  - —with different cultural patterns
- Tests developed in other nations with different language
- Tests developed for international comparisons
- Regional tests developed by OAS or other international agencies
- Local school system tests with an anchor test for comparative outcome analysis

II. Teacher-made Tests

- Use of locally designed teacher-made tests

III. Surveys and Observational Reporting

- Surveys of those participating in program
- Recorded observations of community behavior patterns

FIGURE III-2

STEPS IN CONSTRUCTION OF TESTS

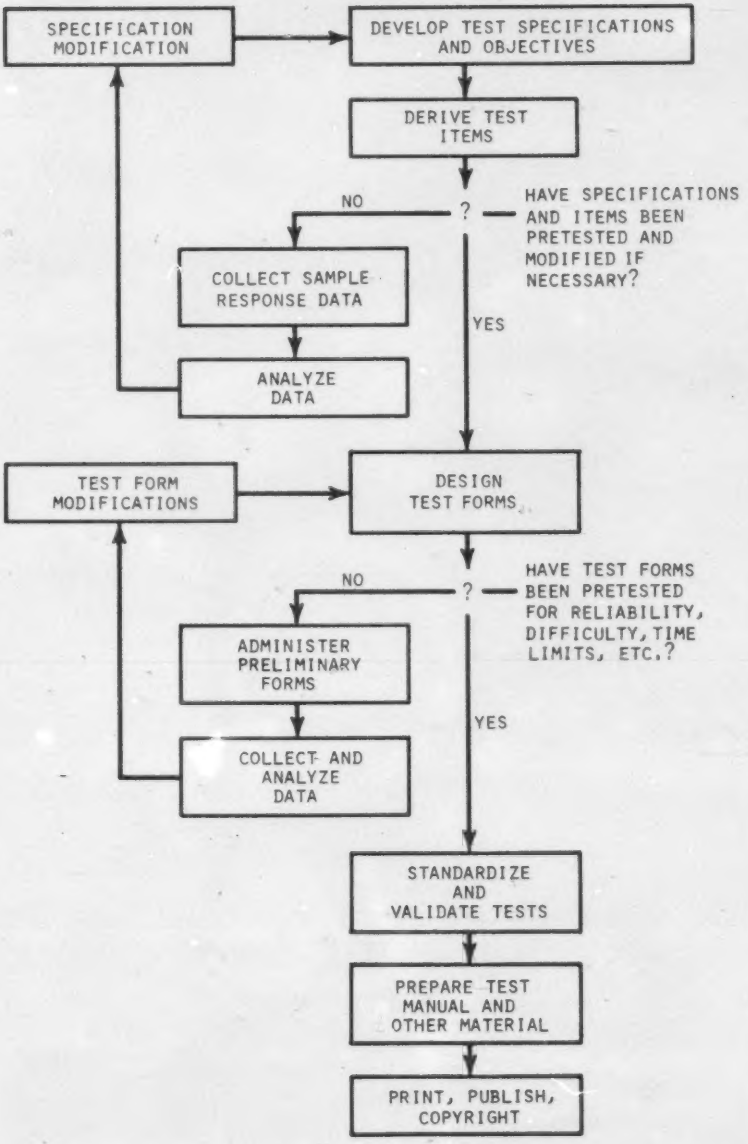


FIGURE III-3

LEVELS OF ADJUSTMENT FOR ADAPTATION OF EXAMINATIONS AND SURVEY INSTRUMENTS

(1) Translation of language.

(2) Changing the vocabulary for the proper reading level.

(3) Changing illustrations and photographs to reflect local conditions

(4) Adjusting procedures implied or specified to match the expectations and experience of the learners.

(5) Adjusting the content to reflect the local culture and life style.

(6) Accommodating the learning styles of the students.

FIGURE III-4

RELATIVE CHARACTERISTICS OF STANDARDIZED TEST

	Validity	*Content Appropriateness*
Standardized Tests	Statistical and empirical evidence of validity	keyed to national or regional goals and priorities
Teacher Evaluation	no direct evidence of validity	keyed to instructional unit in the classroom

ND TEACHER EVALUATION

Uses	*Advantages*	*Disadvantages*
• planning • comparative outcome evaluation of regions and schools • prediction of later school success • feedback to teachers on knowledge of students • feedback of non-formal education to formal for placement, for credentials, etc.	• evaluate all participants on the same set of criteria • standardized tests used at intervals through school career provide longitudinal information • evaluate knowledge, skills where there are no teachers (non-formal education; informal education)	• provide limited information about a student's skills and knowledge • may control instruction and course content • may not measure knowledge covered in local schools
• comparative outcome evaluation of individual students • feedback to teachers on success of instructional methods	• grades are based on cumulative record of performance • show typical behavior and performance in daily classroom situation • provide teacher and student with periodic assessment of achievement • may include assessment of behavior that cannot be measured by tests of skill knowledge	• short-range cognitive and affective changes are not recorded over the school career; cumulative records may not substitute for longitudinal records built up by standardized tests • often compare students within a class instead of providing data on how much has been learned • may be more influenced by good conduct, docility, neatness than on how much a student knows • pressure from parent or community may influence grades

FIGURE III-5A

SELF-CONCEPT AND LOCUS OF CONTROL QUESTIONS*

(1) **How bright do you think you are in comparison with the other students in your grade?**

Agree or disagree:

(2) *"I sometimes feel that I just can't learn."*

(3) *"I would do better in school work if teachers didn't go so fast."*

(4) *"Good luck is more important than hard work for success."*

(5) *"Every time I try to get ahead, something or somebody stops me."*

(6) *"People like me don't have much of a chance to be successful in life."*

Answered either like me or unlike me:

(7) *"I often wish I were someone else."*

(8) *"If I have something to say, I usually say it."*

(9) *"I often get discouraged at what I am doing."*

*The first six questions are from Coleman Report (see Seminar III) and the remaining three are additional sample test items on self-esteem.

FIGURE III-5B

ITEMS ON STUDENT ATTITUDES TOWARD FARM WORK

- *"I like working on the farm because I enjoy it."*
- *"I do not mind doing gardening if its earns me money."*
- *"I like gardening as a hobby."*
- *"I like working on the farm because I have company there."*
- *"If I could choose I would do more interesting things than farming."*

SOURCE: *African Regional Seminar on Education Evaluation.* International Institute for Educational Planning and International Bank for Reconstruction and Development. Dar es Salaam, Tanzania: April/May 1975.